

## 基于规则提示的知识图谱通用推理预训练模型

崔员宁 孙泽群 胡 伟

(计算机软件新技术全国重点实验室(南京大学) 南京 210023)

([yncui.nju@gmail.com](mailto:yncui.nju@gmail.com))

## A Pre-trained Universal Knowledge Graph Reasoning Model Based on Rule Prompts

Cui Yuanning, Sun Zequn, and Hu Wei

(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023)

**Abstract** A knowledge graph (KG) is a structured knowledge base that stores a massive amount of real-world knowledge, providing data support for numerous knowledge-driven downstream tasks. KGs often suffer from incompleteness, with many missing facts. Therefore, the KG reasoning task aims to infer new conclusions based on known facts to complete the KG. With the research and development of knowledge engineering and its commercial applications, numerous general and domain-specific KGs have been constructed. Existing KG reasoning models mostly focus on completing a single KG but lack general reasoning capabilities. Inspired by the general capabilities of pre-trained large language models in recent years, some pre-trained universal KG reasoning models have been proposed. Addressing the issue of existing pre-trained model being unable to identify high-quality reasoning patterns, we introduce a rule-based pre-trained universal KG reasoning model called RulePreM. It discovers and filters high-quality reasoning rules to enhance the reasoning abilities. The proposed model first constructs a relational IO graph based on reasoning rules and uses an encoder, RuleGNN, to encode the relations. The encoded relations are then used as prompts to encode entities in the KG. Finally, candidate entities are scored for prediction. Additionally, an attention mechanism that combines rule confidence is introduced to further reduce the impact of low-quality reasoning patterns. Experimental results demonstrate that the proposed model exhibits universal reasoning abilities on 43 different KGs, with average performance surpassing existing supervised and pre-trained models.

**Key words** knowledge graph; rules; universal reasoning; pre-training; prompt learning; relational IO graph

**摘要** 知识图谱是存储真实世界海量知识的图数据库,为大量知识驱动的下游任务提供了数据支持。知识图谱往往具有不完备性,存在大量缺失的事实,因此知识图谱推理任务基于已知事实推理新结论来补全知识图谱。随着知识工程及其商业应用的研究与发展,大量通用和领域知识图谱被构建。现有知识图谱推理方法大多面向单一知识图谱的补全,不具备通用推理能力。近年来,受预训练大语言模型通用能力的启发,一些通用的知识图谱推理预训练模型被提出。针对现有预训练模型无法识别高质量推理模式的问题,提出一个基于规则提示的知识图谱通用推理预训练模型——RulePreM,该模型筛选与利用高质量推理规则来提高知识图谱上的推理能力。首先基于推理规则构建关系 IO 图和一个编码器 RuleGNN 对关系进行编码,然后将关系编码作为提示来编码知识图谱中的实体,最后对候选实体进行打分预测。还提出一种结合规则置信度的注意力机制,来进一步减少低质量推理模式的影响。实验结果表明,所提出的模型在 43 个不同设定下的知识图谱上具有良好的通用推理能力,平均性能指标均优于现有的有监督模型和预训

收稿日期: 2024-03-15; 修回日期: 2024-04-26

基金项目: 国家自然科学基金项目(62272219)

This work was supported by the National Natural Science Foundation of China (62272219).

通信作者: 胡伟([whu@nju.edu.cn](mailto:whu@nju.edu.cn))

练模型。

**关键词** 知识图谱; 规则; 通用推理; 预训练; 提示学习; 关系 IO 图

**中图法分类号** TP391

知识图谱(knowledge graph, KG)是人类结构化知识的集合<sup>[1-4]</sup>, 存储了从真实世界中抽取的事实型知识。一个知识图谱可以被视作是由实体结点和实体间有向边组成的一个多关系有向图, 其中每个事实被表示为一条实体结点间的有向边(主体实体, 关系, 客体实体)。近年来, 随着知识工程及其应用的快速发展, 大量知识图谱被构建, 如 DBpedia<sup>[5]</sup>, Wikidata<sup>[6]</sup>, 支持了知识库问答、推荐系统、知识检索等下游任务, 满足了不同场景和领域的应用需求<sup>[1-4,7-11]</sup>。

知识图谱推理旨在基于已有事实推理得到新事实, 是支持知识图谱构建与应用的重要任务之一。现有知识图谱推理方法大多面向单一知识图谱, 将每个实体和关系表示为一个可学习的向量, 并基于向量运算来推理得到新事实。不同知识图谱的实体和关系集彼此独立, 这些方法将每个知识图谱视为一个封闭的数据孤岛, 缺乏在不同知识图谱间进行迁移与泛化的能力, 无法进行面向多个知识图谱的通用推理。

最新一些研究受到 GPT<sup>[12]</sup> 等预训练语言模型的启发, 提出了具有知识图谱通用推理能力的预训练模型。文献[13]通过预标注的实体对齐合并不同知识图谱的实体集合, 将多个知识图谱转换为一个大型综合图谱, 提出一种预训练模型进行统一推理。但其知识迁移过程依赖于成本昂贵的预标注实体对齐, 这限制了其在标注数据缺乏等场景下的落地应用。文献[14]提出一种基于相对表示学习的预训练模型 ULTRA, 首先从知识图谱中提取关系图来编码关系向量, 然后基于关系向量来编码实体向量, 避免了对

实体对齐的依赖。尽管如此, 该模型基于简单的邻接关系图来描述推理模式, 无法有效识别和利用高质量的推理模式。如图 1(b)所示, 基于邻接的关系图无法识别和避免低质量的推理路径, 如推理“fatherOf”相关事实时, 存在包含干扰关系“worksFor”的低质量路径: “worksFor (inverse)  $\wedge$  worksFor”, 其中“worksFor (inverse)”表示“worksFor”的逆关系。如图 1(c)所示, 通过引入推理规则及其置信度来选择高质量推理模式, 减少上述干扰关系带来的影响。

本文提出一种基于规则提示的知识图谱通用推理模型, 并基于规则置信度来筛选和利用高质量的知识图谱推理模式。推理规则以待推理关系为规则头、以关系链为规则体, 并以置信度作为规则质量的评价指标。如何利用符号化规则和置信度进行知识图谱通用推理面临 2 项挑战: 一方面, 规则体以关系链的形式存在, 而一个推理关系可能存在多条关系链, 如何融合多条链式规则以及综合利用多条规则中描述的复杂推理模式带来挑战; 另一方面, 如何将数值化的规则置信度结合到神经网络的编码过程中, 从而识别和利用高质量推理模式, 减少低质量推理规则的影响也是一项挑战。

针对综合利用多条规则的挑战, 提出一种关系 IO 图来建模与融合推理规则中复杂的推理模式, 并提出一种关系 IO 图编码器 RuleGNN, 将复杂的关系 IO 图编码为统一的关系表示向量。具体来说, 它的执行包括 3 个阶段: 1) 规则预处理。首先合并规则体中相同的关系, 得到一个简单的关系图; 然后为每个关系设置 1 个输入结点和 1 个输出结点来描述关系间

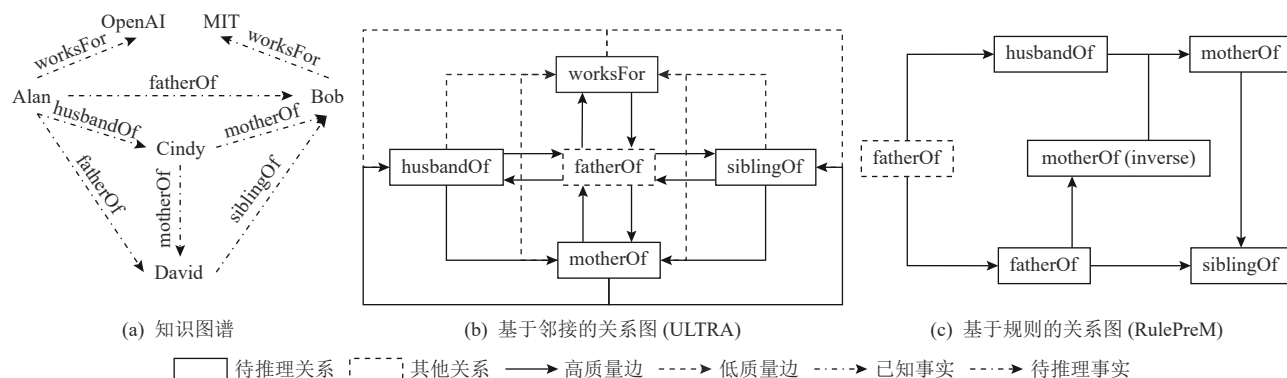


Fig. 1 Comparison of relation graph extracted by ULTRA and RulePreM

图 1 ULTRA 与 RulePreM 提取的关系图对比

的方向信息;最终得到一个综合所有规则的关系 IO 图. 2) 输入向量初始化. 对于结点向量初始化, 为待推理关系的输入结点和输出结点分别设置一个可学习的向量, 其他结点以零向量初始化; 类似地, 对于关系向量初始化, 为待推理关系设置一个可学习的向量, 其它关系以零向量初始化. 3) 图编码. 提出一种基于图神经网络的编码器 RuleGNN, 通过多层迭代在关系 IO 图上传播和聚合来自不同规则的信息, 最终输出统一的关系编码表示向量. 另外, 针对如何利用置信度识别高质量推理模式的挑战, 首先设置了置信度阈值来过滤低质量的规则, 减少关系 IO 图中的低质量规则, 然后提出一种结合置信度的注意力机制, 在 RuleGNN 的信息传播和聚合中进一步降低低质量推理模式带来的影响.

基于所提关系 IO 图和 RuleGNN 对关系的建模与编码, 提出一种基于规则提示的知识图谱推理预训练模型 RulePreM. 该模型将 RuleGNN 编码得到的关系向量作为提示, 在知识图谱上对候选实体进行编码, 并对候选实体进行打分和排序. 为了验证 RulePreM 的通用推理能力, 本文在 3 种设定下的 43 个知识图谱上做了大量实验, 结果表明 RulePreM 的推理性能优于现有的有监督模型和预训练模型.

本文主要贡献包括 4 个方面:

1) 提出了一种基于规则提示的知识图谱推理预训练模型 RulePreM, 该模型基于规则筛选和利用高质量推理模式来增强知识图谱的通用推理能力.

2) 提出一种基于规则的关系 IO 图来融合和利用一个关系相关的多条规则, 并基于规则置信度阈值来筛选高质量的推理模式.

3) 提出一个关系 IO 图编码器 RuleGNN 来聚合和编码不同规则中的关系路径信息, 并提出一种结合置信度的注意力机制来减少低质量推理模式的影响.

4) 多种设定下的 43 个数据集上的实验结果表明, RulePreM 具有良好的通用推理能力, 所提关系 IO 图及其编码为预训练模型提供了有效提示.

## 1 相关工作

### 1.1 知识图谱推理模型

由于知识源和信息抽取的局限性, 现有知识图谱中往往存在事实大量缺失. 知识图谱的不完备性是知识工程中面临的重要挑战, 它限制了下游应用的研究与发展. 因此, 知识图谱推理任务基于已有事实推理得到的新事实来补全知识图谱.

知识图谱推理主要包含 3 种设定: 归纳式推理、全归纳式推理和转导式推理.

1) 归纳式推理 (inductive reasoning)<sup>[15]</sup> 假设知识图谱中的关系集是固定的, 而实体和事实是在不断变化的. 归纳式推理设定下的主要模型包括基于邻居实体聚合的模型<sup>[16-18]</sup>、基于子图归纳的模型<sup>[19-22]</sup>和基于相对表示学习的模型<sup>[23-26]</sup>. 其中基于邻居实体聚合的模型如 LAN<sup>[17]</sup> 通过聚合新实体的已知邻居实体来表示新实体, 无法处理与已知实体无直接关联的实体; 基于子图归纳的模型以 GraIL<sup>[19]</sup> 为代表, 抽取主体和客体实体之间的子图来为候选实体打分, 它们要为每个候选实体抽取子图, 无法用于大规模知识图谱; 以 NBFNet<sup>[23]</sup> 为代表的基于相对表示学习的模型以待测事实为条件, 将待测事实嵌入到实体编码, 具有良好的性能和效率.

2) 全归纳式推理 (fully-inductive reasoning)<sup>[27-30]</sup> 假设知识图谱中的实体和关系集合都是动态变化的, 较归纳式推理更具挑战性. 代表性工作如 InGram<sup>[27]</sup>, 以归纳式的或随机动态的表示作为各关系的初始表示, 使模型能在训练过程中学习基于随机的特征来表征图结构特征, 因此能够自然地适配到新关系上. 尽管这些方法能够处理新实体和新关系, 但它们并不能处理新知识图谱.

3) 转导式推理 (transductive reasoning)<sup>[31-36]</sup> 假设知识图谱是静态的, 如代表性工作 Rotate<sup>[35]</sup>, 将实体和关系编码为表示向量, 然后基于向量计算来补全不完整的事实. 近年来一些基于卷积神经网络<sup>[33]</sup> 和图神经网络<sup>[36]</sup> 的工作也陆续被提出, 但它们仍然在第一层输入中将每个实体和关系映射到特定的表示向量, 无法处理新增实体和关系. 为验证所提出的模型 RulePreM 的有效性, 本文在上述 3 种设定下的 43 个数据集上开展大量实验. 结果表明 RulePreM 在各种设定下的数据集上具有良好的通用推理能力.

另外, 除了上述的基于参数化表示学习的模型, 基于规则的模型也是一类知识图谱推理模型. 首先从知识图谱的训练集中挖掘推理规则, 并基于置信度等量化指标来计算每条推理规则的质量; 然后基于提取的规则来预测缺失的事实和补全知识图谱. 早期的基于规则的模型, 如 AMIE<sup>[37]</sup>, 主要关注如何基于剪枝和并行技术来提升规则挖掘效率. NeurLP<sup>[38]</sup> 和 DRUM<sup>[39]</sup> 采用了端到端的差分学习方式, 通过最大化所有事实分数来挖掘规则. AnyBURL<sup>[40]</sup> 扩展了规则的定义, 自底向上地利用上下文信息挖掘规则, 是一种效率很高的规则挖掘和置信度估计模型.



RulePreM 在规则预挖掘模块也采用了 AnyBURL 100s 的版本. 尽管基于规则的模型具有可解释等良好特性, 但它们在推理时不能考虑知识图谱除路径外的图上信息, 因此基于规则的模型推理性能受限. 基于规则及其置信度, 针对现有的知识图谱预训练模型无法鉴别高质量推理模式的问题, 提出一种筛选和利用高质量推理模式的模型. 具体来说, 将规则用于关系编码, 降低了来自低质量推理模式的信息的影响.

## 1.2 图预训练模型

与所提模型 RulePreM 相关的另一类工作是图的预训练模型. 在自然语言处理领域, BERT<sup>[41]</sup> 和 GPT3<sup>[12]</sup> 等预训练模型的提出深刻影响了各类下游任务. 它们基于迁移学习的范式, 首先在源数据集上预训练一个模型, 然后迁移到目标数据集或任务上, 通过微调或提示等手段使预训练模型能够快速适应新数据集或任务. 在这样的背景下, 一些遵循“预训练-微调”的范式的图预训练模型<sup>[42-46]</sup> 被提出. 如 GraphMAE<sup>[42]</sup> 提出一种掩码自编码图预训练模型, 先在预训练数据集上做自监督预训练, 然后在各下游任务上微调. 然而, 这类预训练模型难以处理下游任务所需知识的差异性, 有时预训练不仅不会带来有效知识迁移, 还会降低下游任务性能. 针对下游任务的差异性, 一些基于“预训练-提示-微调”的图预训练模型<sup>[47-56]</sup> 被提出. 该模型设计提示图编码来激发预训练模型与下游任务相关的知识迁移, 从而提升下游任务性能. 如 AllInOne<sup>[54]</sup> 设计了一个额外的可学习提示图来表征不同的下游任务, 实现了更好的知识迁移. 所提出的基于规则提示的预训练模型也遵循了这一范式.

另外, 大多数现有知识图谱推理模型仅能在一个知识图谱上应用. 近年来, 知识图谱领域一些最新工作探索了基于知识图谱的预训练. 为了打破知识图谱之间的数据孤岛, 一些工作<sup>[13,57-60]</sup> 首先基于实体对齐<sup>[61-67]</sup> 来合并不同知识图谱的实体集, 然后将预训练模型在待测知识图谱上微调. 然而, 不是所有的知识图谱间都有实体对齐标注, 昂贵的预标注成本限制了其应用场景. ULTRA<sup>[14]</sup> 提出了一个不需要实体对齐的预训练推理模型, 它基于关系图编码关系向量和执行推理. 针对其关系图无法识别和利用高质量推理模式的问题, 引入推理规则和置信度来避免低质量推理模式的影响. 另外, 文献[68] 还提出了一种知识图谱预训练模型 KGTransformer, 它的目标不是知识图谱推理, 而是图像分类等更广泛的下游任务.

## 2 任务定义

本节主要定义知识图谱推理任务相关的一些概念, 以及知识图谱推理规则.

**定义 1.** 知识图谱. 一个知识图谱是一个多关系有向图  $G = (E, R, T)$ .  $E$  代表实体集,  $R$  代表关系集,  $T$  代表事实集. 事实集中的每条事实被表示为一个三元组  $(s, r, o)$ , 其中  $s \in E$  称作主体实体,  $o \in E$  称作客体实体,  $r \in R$  称作 2 个实体之间的关系.

知识图谱是一个有向图, 现有工作大多引入逆关系及其相关事实来扩充知识图谱, 如图 1(c) 中的关系 “motherOf (inverse)” 是关系 “motherOf” 的逆关系. 本文也采用了这一扩充方法, 对于一个关系  $r \in R$ , 将其逆关系  $r^{-}$  引入到关系集  $R$  中, 并将其相关事实  $(s, r, o)$  的反向事实  $(o, r^{-}, s)$  加入到事实集  $T$  中.

知识图谱推理旨在基于已有事实, 推理得到新事实来补全知识图谱中实体间缺失的链接. 定义有:

**定义 2.** 知识图谱推理. 对于给定的知识图谱  $G = (E, R, T)$  和一系列包含缺失实体的形如  $(s, r, x)$  的待推理事实, 知识图谱推理任务以在  $E$  中的实体为候选集, 预测目标实体  $x$  来补全不完整的待推理事实.

基于推理规则来筛选和利用高质量推理模式, 知识图谱中的推理规则定义有:

**定义 3.** 推理规则. 一条推理规则  $l$  由规则头和规则体组成, 形如  $q \leftarrow r_1 \wedge \cdots \wedge r_n$ , 其中规则头  $q \in R$  是待推理事实中包含的待推理关系,  $r_1, \cdots, r_n$  是规则体中的关系. 包含在  $l$  的规则体中的一个关系  $r$  与  $l$  之间的关系记作  $r \in l$ . 每条推理规则有一个置信度  $\text{conf}(l)$ , 表示该规则的可信程度. 一个知识图谱中的规则集合记作  $L$ , 其中以关系  $q$  为规则头的规则的集合记作  $L_q$ .

## 3 本文所提出的模型

本节详细介绍如何基于规则提示来支持通用推理预训练模型. 具体来说, 分为 3 个步骤: 基于规则的关系 IO 图构建、关系 IO 图编码、实体编码与推理.

### 3.1 基于规则的关系 IO 图构建

针对知识图谱预训练模型无法鉴别高质量推理模式的问题, 探索如何借助推理规则来增强预训练模型的通用能力. 因此, 首先基于规则挖掘算法来挖掘规则和获取其置信度.

在过去的知识图谱推理相关的研究工作中, 已有很多知识图谱规则的挖掘算法被提出<sup>[37-39]</sup>. 然而, 大多数规则挖掘算法都非常耗时和消耗计算资源, 这与知识图谱预训练模型对于在目标知识图谱上快速部署的目标相悖. 所提出的模型采用 AnyBURL 100s<sup>[40]</sup> 作为规则挖掘器, 它是一种快速的规则挖掘方法, 仅需 100 s 的时间来挖掘规则. 其挖掘的规则包含 2 种: 一种是仅由关系组成的关系规则; 另一种是包含实体的混合规则, 提取其中的关系规则, 并设置了一个置信度阈值  $\varepsilon$ , 仅保留置信度高于  $\varepsilon$  的高质量规则.

接下来, 基于筛选后的规则来构建关系图, 将具有相同规则头的所有规则合并为一个如图 2 左图所

示的关系图. 首先构建结点集, 对于待推理关系  $q$ , 为  $q$  和每个出现在以  $q$  为规则头的规则体中的关系分别设置 1 个结点. 然后构建关系图的边集. 对于规则  $q \leftarrow r_1 \wedge \dots \wedge r_n$ , 首先添加规则头  $q$  与每个规则体中的第 1 个关系之间的边  $(q, r_1)$ , 然后添加规则体中前序关系指向下一个关系的边  $(r_i, r_{i+1})$ . 最后, 取所有边集的合集作为  $q$  的关系图的边集.

图 2 左图所示的关系图仍然存在 2 个问题: 一是它缺少同一个关系和其逆关系之间的边, 这将会造成信息损失; 二是它对关系结点间方向建模不够充分. 因此进一步扩展关系图, 提出了一种能更好地描述复杂推理模式的关系 IO 图, 其定义有:

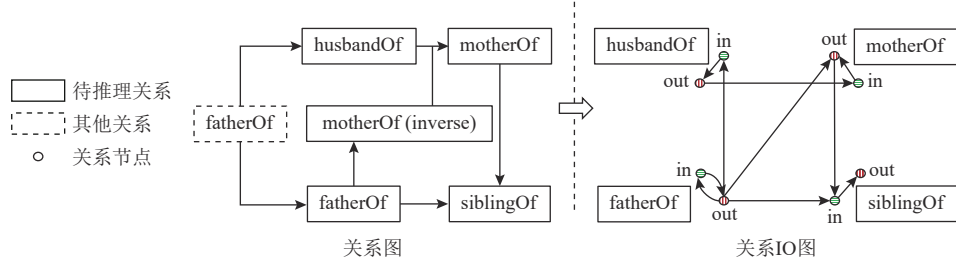


Fig. 2 An example of rule-based relational IO graph construction

图 2 基于规则的关系 IO 图构造实例

**定义 4. 关系 IO 图.** 关系 IO 图用于描述规则中存在的关系之间的连接关系, 是一个多关系有向图  $M_q = (V_{\text{rule}}, R_{\text{rule}}, D_{\text{rule}})$ . 其中  $q$  是关系 IO 图中规则的规则头对应的待推理关系;  $V_{\text{rule}}$  为结点集, 它包含每一个关系  $r$  的一个输入结点  $v_r^{\text{in}}$  和一个输出结点  $v_r^{\text{out}}$ ;  $R_{\text{rule}}$  是关系集;  $D_{\text{rule}}$  是边集, 每条边形如  $(n_x, r, n_y)$ , 其中  $n_x, n_y \in V_{\text{rule}}$  是 2 个结点,  $r$  是边的关系类型. 每条边对应 1 个置信度值, 置信度的值等于包含该关系的规则置信度最大值, 即  $\text{conf}(n_x, r, n_y) = \max\{\text{conf}(l) | \forall r \in l\}$ .

为每个关系  $r$  设置一个输入结点和一个输出结点来表征规则中关系的方向. 其中  $r$  的逆关系  $r^{-1}$  的输入结点和  $r$  的输出结点是同一个结点, 即  $v_r^{\text{in}} = v_{r^{-1}}^{\text{out}}$ , 同理,  $v_r^{\text{out}} = v_{r^{-1}}^{\text{in}}$ . 然后添加关系 IO 图的边集, 它包含 2 部分: 第 1 部分是同一个关系的输入结点和输出结点之间的边  $(v_r^{\text{in}}, r, v_r^{\text{out}})$ ; 第 2 部分是规则中相邻关系之间的边, 如规则体中包含的 2 个相邻关系  $r_i$  和  $r_{i+1}$ , 添加边  $(v_{r_i}^{\text{out}}, r_i, v_{r_{i+1}}^{\text{in}})$ . 与关系图相比, 关系 IO 图用输入和输出结点更好地建模了规则中关系间的方向信息, 并提供了关系与其逆关系之间信息传递的通道.

### 3.2 基于关系 IO 图的关系编码

本节介绍如何对关系 IO 图进行编码. 关系 IO 图是一个多关系有向图, 并提出了一个规则编码器

RuleGNN 来执行编码. 其编码过程包含 2 个步骤: 初始化输入向量表示和关系编码.

首先, 为关系 IO 图中的结点和关系初始化输入表示向量. 由于预训练知识图谱推理模型不止面向单一知识图谱, 而是要在多个知识图谱上具有通用能力, 所以初始化表示向量中不能包含特定知识图谱关系的相关参数. 因此, 对于一个待推理事实  $(s, q, x)$ , 只为待推理关系  $q$  设置了一个随机初始化的可学习向量, 其他关系以零向量作为初始化输入向量. 其形式化定义有:

$$r^{(0)} = \begin{cases} q, & \text{if } r = q; \\ \mathbf{0}, & \text{其他,} \end{cases} \quad (1)$$

其中  $q \in \mathbb{R}^{1 \times d}$  表示一个随机初始化可学习向量,  $d$  是表示向量的维度,  $\mathbf{0}$  表示零向量. 以类似的方式来初始化关系 IO 图中的结点的输入表示向量:

$$v^{(0)} = \begin{cases} q^{\text{in}}, & \text{if } v = v_q^{\text{in}}; \\ q^{\text{out}}, & \text{if } v = v_q^{\text{out}}; \\ \mathbf{0}, & \text{其他,} \end{cases} \quad (2)$$

其中  $q^{\text{in}} \in \mathbb{R}^{1 \times d}$  和  $q^{\text{out}} \in \mathbb{R}^{1 \times d}$  是 2 个随机初始化的可学习向量. 基于式 (1)(2), 为每个结点和关系都设置了一个输入向量, 同时避免了针对特定知识图谱的参数.

然后,提出一种图编码器 RuleGNN 在关系 IO 图上传递消息和获取关系表示向量. RuleGNN 的编码过程包括结点编码更新和关系编码更新 2 个步骤. 首先,聚合一个结点  $v$  相关边集中的消息来更新中心结点的表示向量:

$$\mathbf{v}^{(i+1)} = \frac{1}{|D_v|} \sum_{(u,r,v) \in D_v} \alpha \mathbf{W}_1^{(i)}(\mathbf{u}^{(i)}; \mathbf{r}^{(i)}), \quad (3)$$

$$\alpha = \text{conf}(u, r, v) + \text{Sigmoid}(\mathbf{W}_2^{(i)}(\mathbf{r}^{(i)}; \mathbf{q}^{(i)})),$$

其中  $D_v$  为包含中心结点  $v$  的相关事实集,  $\mathbf{W}_1^{(i)} \in \mathbb{R}^{d \times 2d}$  和  $\mathbf{W}_2^{(i)} \in \mathbb{R}^{d \times 2d}$  是 2 个可学习的权重矩阵,  $\mathbf{u}^{(i)}$ ,  $\mathbf{v}^{(i)}$ ,  $\mathbf{r}^{(i)}$  和  $\mathbf{q}^{(i)}$  分别表示结点  $u$ 、结点  $v$ 、关系  $r$  和待推理关系  $q$  在第  $i$  层的表示向量.  $\alpha$  是该边的注意力权重,其由该边的置信度和由神经网络计算的注意力权重相加得到.  $(;)$  表示拼接操作,  $\text{Sigmoid}$  表示激活函数.

接下来, RuleGNN 更新关系 IO 图中的关系表示向量. 对于一个关系  $r$ , 聚合与其相关的边集中的消息来更新它下一层的表示向量:

$$\mathbf{r}^{(i+1)} = \frac{1}{|D_r|} \sum_{(u,r,v) \in D_r} \beta \mathbf{W}_3^{(i)}(\mathbf{u}^{(i+1)}; \mathbf{v}^{(i+1)}), \quad (4)$$

$$\beta = \text{conf}(u, r, v) + \text{Sigmoid}(\mathbf{W}_4^{(i)}(\mathbf{u}^{(i+1)}; \mathbf{v}^{(i+1)}; \mathbf{q}^{(i)})),$$

其中  $D_r$  为包含中心结点  $r$  的相关事实集,  $\mathbf{W}_3^{(i)} \in \mathbb{R}^{d \times 2d}$  和  $\mathbf{W}_4^{(i)} \in \mathbb{R}^{d \times 3d}$  是 2 个可学习的权重矩阵.  $\beta$  是一个由该边的置信度和神经网络计算的权重相加得到的注意力权重. 注意, 尽管边的置信度由其中包含的关系决定, 式(4)中同一个关系的相关边置信度都是相同的, 但将式(4)用于不同的关系时, 置信度仍然能起到对不同关系的表示向量缩放的作用, 因此仍然将边的置信度作为注意力权重的一部分.

最后, 读取最后一层关系作为关系 IO 图输出的关系编码, 即  $\mathbf{r}^{\text{rule}} = \mathbf{r}^{(L)}$ , 其中  $\mathbf{r}^{\text{rule}}$  是关系  $r$  经过 RuleGNN 编码后的输出向量,  $L$  是一个超参数, 表示 RuleGNN 的预设层数. RuleGNN 没有与特定知识图谱绑定的可学习参数, 因此保留了通用推理能力.

### 3.3 基于知识图谱的实体编码

基于关系 IO 图和 RuleGNN 提供的关系编码, 编码知识图谱中的候选实体并打分. 传统的知识图谱编码大多为每个实体和关系都设定了一个可学习的表示向量, 这不能满足预训练模型对于通用性的要求. 因此, 受归纳式知识图谱表示学习方法 RED-GNN<sup>[24]</sup> 的启发, 采用相对知识图谱编码的方式, 基于关系编码来得到实体的编码.

具体来说, 对于一个关系  $r$ , 首先基于 RuleGNN 得到的最终关系编码来赋值输入表示向量:

$$\mathbf{r}^{(0)} = \begin{cases} \mathbf{W}_5 \mathbf{r}^{\text{rule}}, & \text{if } r \in D_{\text{rule}}; \\ \mathbf{0}, & \text{其他,} \end{cases} \quad (5)$$

其中  $\mathbf{W}_5 \in \mathbb{R}^{d \times d}$  是一个可学习的权重矩阵, 用于对从关系 IO 图中读取出来的关系编码的转换. 对于不存在于关系 IO 图中的关系, 用零向量作为表示向量, 以降低对推理产生干扰关系的影响.

对于一个待推理事实  $(s, q, x)$ , 以关系  $q$  对应的关系表示向量, 来编码知识图谱中的候选实体. 借鉴 RED-GNN<sup>[24]</sup> 采用的逐跳扩展候选实体的方法, 以主体实体  $s$  为根节点, 逐跳扩展需要编码的候选实体, 来充分利用相对距离信息来对候选实体编码. 具体来说, 对于一个实体  $e$ , 用包含它的事实集合更新它下一层的表示向量:

$$\mathbf{e}^{(i+1)} = \mathbf{W}_6^{(i)} \mathbf{e}^{(i)} + \frac{1}{|T_e^{(i+1)}|} \sum_{(u,r,e) \in T_e} \lambda \mathbf{W}_7^{(i)}(\mathbf{u}^{(i)} + \mathbf{r}^{(i)}), \quad (6)$$

$$\lambda = \text{Sigmoid}(\mathbf{W}_8^{(i)}(\mathbf{u}^{(i)}; \mathbf{r}^{(i)}; \mathbf{q}^{(i)})),$$

其中  $\mathbf{u}^{(i)}$ ,  $\mathbf{r}^{(i)}$  和  $\mathbf{q}^{(i)}$  是实体  $u$ 、关系  $r$  和待推理关系  $q$  在第  $i$  层的表示向量,  $\lambda$  是一个用神经网络计算的注意力权重,  $\mathbf{W}_6^{(i)} \in \mathbb{R}^{d \times d}$ ,  $\mathbf{W}_7^{(i)} \in \mathbb{R}^{d \times d}$  和  $\mathbf{W}_8^{(i)} \in \mathbb{R}^{d \times 2d}$  是 3 个可学习的权重矩阵.  $T_e^{(i)}$  是  $s$  的  $i$  跳邻域内包含实体  $e$  的事实集合, 将  $T_e^{(i)}$  形式化表示为:

$$\begin{aligned} E^{(0)} &= \{s\}, \\ E^{(i+1)} &= \{e | \exists (x, r, e) \in T \wedge x \in E^{(i)}\}, \\ T^{(i)} &= \{(x, r, y) | x \in E^{(i)} \wedge y \in E^{(i)} \wedge (x, r, y) \in T\}, \end{aligned} \quad (7)$$

其中  $E^{(i)}$  表示待推理事实的主体实体  $s$  的第  $i$  跳邻域内包含实体的集合. 随着层数的增加, 模型所编码的实体集不断扩展, 因此它能够利用各候选实体与待推理事实中主体实体  $s$  之间的距离信息.

在每一层的最后更新下一层的关系表示, 以更好地表征每一层中关系不同的语义:

$$\mathbf{r}^{(i+1)} = \mathbf{W}_9^{(i)} \mathbf{r}^{(i)}, \quad (8)$$

其中  $\mathbf{W}_9^{(i)} \in \mathbb{R}^{d \times d}$  是一个可学习的权重矩阵.

在多层迭代更新后, 对于每个实体  $e$ , 读取其在最后一层更新得到的表示向量  $\mathbf{e}^{\text{KG}} = \mathbf{e}^{(H)}$  作为输出向量, 其中  $H$  是一个预设的超参数, 表示知识图谱编码器的层数.

### 3.4 模型预训练优化目标

给定用于预训练的知识图谱列表  $P = \{G_1, \dots, G_{|P|}\}$ , 设计了损失函数:

$$\mathcal{L} = \sum_{G_i \in P} \sum_{(s,q,o) \in T_i} \left( -\text{score}(s, q, o) + \log \left( \sum_{e \in E_i} \exp(\text{score}(s, q, e)) \right) \right), \quad (9)$$

其中  $\text{score}()$  是一个评分函数. 基于从知识图谱编码



模块读取的实体编码设计了评分函数:

$$\text{score}(e) = \mathbf{W}_{10} \mathbf{e}^{\text{KG}}, \quad (10)$$

其中  $\mathbf{W}_{10} \in \mathbb{R}^{d \times d}$  是一个可学习的权重矩阵. 对于不在  $s$  的  $W$  跳以内的其他实体, 其分数设为 0.

### 3.5 预训练与推理流程

本节对 3.1~3.4 节描述的步骤做汇总, 来更清晰和直观地展示模型如何运行. 具体来说, 模型的执行过程包括预训练阶段和推理阶段. 在预训练阶段中, 模型基于预训练数据集中的训练样本计算损失函数和更新模型参数. 在推理阶段, 预训练模型的参数将被冻结, 仅在待推理数据集上执行评分和推理.

具体的预训练算法如算法 1 所示.

**算法 1.** 知识图谱推理预训练算法.

输入: 预训练知识图谱集合  $P = \{G_1, \dots, G_{|P|}\}$ , 随机初始化的模型参数  $\theta$ , 预设最大迭代数  $\text{maxepoch}$ , 预设早停步数  $ms$ ;

输出: 更新后的模型参数  $\theta$ .

```

/*数据预处理*/
① 获取每个知识图谱的规则集合  $L_1, \dots, L_{|P|}$ ;
② 生成每个关系  $q$  的关系 IO 图  $M_q$ ;
/*模型训练*/
③ for epoch=1, ..., maxepoch
④   for each sample( $s, q, o$ )
/*关系编码, 3.2 节*/
⑤      $\mathbf{R}^{\text{rule}} \leftarrow \text{RuleGNN}(M_q)$ ;
/*实体编码, 3.3 节*/
⑥      $\mathbf{E}^{\text{kg}} \leftarrow \text{KGEncoder}(\mathbf{R}^{\text{rule}}, G_s)$ ;
/*模型优化, 3.4 节*/
⑦     式 (10) 计算候选实体分数;
⑧     式 (9) 计算损失;
⑨     反向传播更新模型参数  $\theta$ ;
⑩   end for
⑪   result  $\leftarrow \text{eval}(\text{validation set})$ ;
⑫   if 累积  $ms$  次结果未比历史最优结果高
⑬     break;
⑭   end if
⑮ end for

```

算法 1 中  $\mathbf{R}^{\text{rule}}$  表示经过 RuleGNN 编码后得到的关系向量表示的矩阵,  $\mathbf{E}^{\text{kg}}$  表示经过知识图谱编码后得到的实体向量表示的矩阵, KGEncoder 表示 4.3 节介绍的知识图谱编码器,  $G_s$  表示该训练样本所在的知识图谱.

在推理阶段, 首先基于算法 1 中的行①~②做预处理. 随后执行算法 1 中行④~⑩所示的步骤, 先基

于 RuleGNN 对关系 IO 图编码得到关系表示向量, 然后基于知识图谱编码器得到实体表示向量, 最后对所有候选实体打分和排序来得到推理结果.

## 4 实验结果与分析

### 4.1 数据集

为了验证所提模型 RulePreM 的有效性, 在通用的 3 种知识推理设定下的 43 个知识图谱上进行了知识图谱推理实验:

1) 归纳式推理设定. 归纳式推理被用于验证模型在具有新实体的动态知识图谱上的推理能力, 在归纳式推理数据集中, 测试集和训练集在 2 个不同的子图上, 测试集中包含训练集中未出现的实体. 具体来说, 实验采用了 14 个归纳式数据集, 其中包含 12 个来自 GraIL<sup>[19]</sup> 的数据集和 2 个来自 ILPC<sup>[69]</sup> 的数据集.

2) 全归纳式推理设定. 全归纳式推理设定被用于验证模型在具有新实体和新关系的动态知识图谱上的推理能力, 在全归纳式推理数据集中, 测试集中可能包含训练集中未出现的实体和关系. 具体来说, 实验采用了 13 个来自文献 [27] 的全归纳式推理数据集.

3) 转导式推理设定. 转导式推理设定被用于验证模型在静态知识图谱上的推理能力. 在转导式推理数据集中, 测试集包含的实体和关系也都包含在训练集中. 具体来说, 实验采用了 16 个转导式推理数据集, 包括 FB15k-237<sup>[70]</sup>, WN18RR<sup>[33]</sup>, NELL-995<sup>[71]</sup>, YAGO3-10<sup>[72]</sup>, 3 个来自 CoDex<sup>[73]</sup> 的数据集, 5 个来自文献 [74] 的数据集, AristoV4<sup>[75]</sup>, DBpedia100k<sup>[76]</sup>, ConceptNet100k<sup>[77]</sup> 和 Hetionet<sup>[78]</sup>.

### 4.2 实验设定

1) 评价指标. 对于测试集中的每条事实  $(s, r, o)$ , 依次将主体实体和客体实体作为待预测的目标实体, 构建 2 个待推理事实, 分别为  $(s, r, x)$  和  $(o, r, x)$ . 然后将数据集的所有实体作为候选实体, 模型为每个候选实体打分, 将所有的候选实体按照分数进行排序. 采用了 2 个知识图谱推理通用的评价指标<sup>[34]</sup> 作为性能评价标准: ① 平均倒数排名 (mean reciprocal rank,  $MRR$ ), 即目标实体的排名的倒数的平均值; ② 预测实体在前 10 的百分比 ( $Hits@10$ ), 越高的指标值表明性能越好. 另外, 沿用知识图谱推理实验中通用的正确实体过滤策略, 对于  $(s, r, x)$  可能有多个目标实体都是正确的情况, 将除了目标实体以外的已知正

确实体从候选实体中移除,这有助于避免正确的预测被打低分数的情况.另外,由于用于测评的数据集数目过多,为了方便展示,将数据集按照 1.1 节中介绍的 3 种设定进行分组,首先计算每个数据集上的指标分数,然后计算并报告组内指标的平均值作为该设定下知识图谱推理性能的平均指标.

2)模型预训练设定.首先介绍超参数设定情况:采用了 5 层的 RuleGNN 作为关系 IO 图编码器和 6 层的知识图谱作为实体编码器.设置表示向量的维度  $d=48$ ,置信度阈值  $\varepsilon=0.3$ ,最大迭代数  $maxepoch=100$ ,早停步数  $ms=5$ .采用 Xavier 作为可学习参数初始化的方法,用 Adam 优化器来更新模型参数,学习率设为 0.001.然后介绍模型的预训练设定:采用 CoDEx-small<sup>[73]</sup>(42 个关系),FB V1<sup>[19]</sup>(180 个关系),和 NELL V1<sup>[19]</sup>(14 个关系)作为预训练数据集.在完成预训练之后,模型参数将被冻结,并直接用于评测数据集上的实验.在一个配有 4 个 NVIDIA RTX 3090 GPU,2 个 Inter Xeon Gold 5122 CPU 和 384 GB 的内存的工作站上预训练模型.预训练所需时长小于 5 h,模型参数量为 207 000.

3)基线模型.基线模型包含 2 类.第 1 类是各数据集上的有监督学习的先进模型(supervised state-of-the-art, SOTA),简称为“有监督 SOTA”,是每个数据集上取得最优 MRR 指标分数的模型.其中归纳式数据集上有监督 SOTA 模型包括 NBFNet<sup>[23]</sup>(在 FB V2<sup>[19]</sup>, FB V3<sup>[19]</sup>, WN V1<sup>[19]</sup>和 WN V3<sup>[19]</sup>数据集上),RED-GNN<sup>[24]</sup>(在 NELL V1, NELL V2, NELL V3 和 NELL V4<sup>[19]</sup>数据集上),A\*Net<sup>[25]</sup>(在 FB V1<sup>[19]</sup>, FB V4<sup>[19]</sup>和 WN V4<sup>[19]</sup>数据集上)和 NodePiece<sup>[18]</sup>(在 ILPC-small<sup>[69]</sup>和 ILPC-large<sup>[69]</sup>数据集上);全归纳式数据集上的有监督 SOTA 模型为 InGram<sup>[27]</sup>;转导式数据集上的有监督 SOTA 模型包括 NBFNet<sup>[23]</sup>(在 FB15k-237<sup>[70]</sup>, WN18RR<sup>[33]</sup>和 YAGO3-10<sup>[72]</sup>数据集上),ComplEx RP<sup>[79]</sup>(在 CoDEx-small<sup>[73]</sup>, CoDEx-medium<sup>[73]</sup>, CoDEx-large<sup>[73]</sup>

和 AristoV4<sup>[75]</sup>数据集上),RED-GNN<sup>[24]</sup>(在 NELL-995<sup>[71]</sup>, RotatE<sup>[35]</sup>和 Hetionet<sup>[78]</sup>数据集上),LR-GCN<sup>[80]</sup>(在 WDSinger<sup>[74]</sup>和 NELL23k<sup>[74]</sup>数据集上),DacKGR<sup>[81]</sup>(在 FB15k-237-10<sup>[74]</sup>, FB15k-237-20<sup>[74]</sup>和 FB15k-237-50<sup>[74]</sup>数据集上),BiQUE<sup>[81]</sup>(在 ConceptNet100k<sup>[77]</sup>数据集上)和 ComplEx-NNE+AER<sup>[76]</sup>(在 DBpedia100k<sup>[76]</sup>数据集上).第 2 类方法是预训练模型 ULTRA<sup>[14]</sup>.ULTRA 提供了 ULTRA-预训练和 ULTRA-微调 2 个版本,其中预训练版本在模型预训练结束后冻结参数;而微调版本在预训练结束后,分别在每个待测数据集上微调,然后用于该待测数据集测试.为了与 ULTRA 进行全方面对比,在主实验中也提供了 RulePreM-预训练和 RulePreM-微调 2 个版本,其中微调版本预训练之后,在待测数据集上再微调 5 轮迭代.在 4.3 节和 4.4 节之外的实验中,基于预训练版本做进一步分析.

### 4.3 主实验结果

主实验采用了 3 种设定下的 43 个知识图谱作为评测数据集,并汇报了 5 次运行的平均结果与标准差.在有监督 SOTA 模型中,一些模型未开源代码,因此仅引用文献 [14] 中报告的结果.各设定数据集的平均指标如表 1 所示. RulePreM 的预训练和微调版本在各类数据集上均取得了良好性能.其中 RulePreM 在归纳式推理和转导式推理数据集的平均分数优于各类基线方法, RulePreM 在全归纳式推理的 13 个数据集上的平均指标,以及在全 43 个数据集上指标的平均值优于基线方法.在归纳式推理数据集和转导式推理数据集上,微调版本的结果都优于预训练版本,而在全归纳式推理数据集上,微调版本的结果有些许下降.这主要是由于在全归纳式推理数据集中,训练图与测试图中的关系集合是不同的,模型未能通过微调获取更好的关系表征.在转导式推理数据集上, RulePreM 的性能相比于有监督 SOTA 的提升比另外 2 种数据集上的低.这是因为转导式推理数据集中的训练集和测试集共享相同的实体和关系集

Table 1 MRR and Hits@10 of the Main Experiment

表 1 主实验的 MRR 和 Hits@10

| 模型           | 归纳式推理 (14 个数据集)    |                    | 全归纳式推理 (13 个数据集)   |                    | 转导式推理 (16 个数据集)    |                    | 平均值 (43 个数据集)      |                    |
|--------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|              | MRR                | Hits@10            | MRR                | Hits@10            | MRR                | Hits@10            | MRR                | Hits@10            |
| 有监督 SOTA     | 0.466              | 0.607              | 0.210              | 0.340              | 0.365              | 0.511              | 0.351              | 0.493              |
| ULTRA-预训练    | 0.512±0.002        | 0.663±0.003        | 0.353±0.005        | 0.536±0.005        | 0.330±0.004        | 0.479±0.004        | 0.396±0.001        | 0.557±0.001        |
| ULTRA-微调     | 0.529±0.005        | 0.684±0.004        | 0.349±0.007        | 0.543±0.005        | <u>0.384±0.005</u> | <u>0.547±0.006</u> | 0.421±0.002        | 0.590±0.001        |
| RulePreM-预训练 | <u>0.550±0.003</u> | <u>0.698±0.003</u> | <b>0.420±0.005</b> | <b>0.621±0.004</b> | 0.331±0.004        | 0.476±0.006        | <u>0.429±0.002</u> | <u>0.592±0.002</u> |
| RulePreM-微调  | <b>0.563±0.004</b> | <b>0.710±0.003</b> | <u>0.419±0.006</u> | <u>0.619±0.005</u> | <b>0.387±0.004</b> | <b>0.549±0.003</b> | <b>0.454±0.001</b> | <b>0.622±0.001</b> |

注: 黑体数值表示最优值; 下划线数值表示次优值.



合,这使得有监督 SOTA 模型可以在训练阶段充分学习实体和关系表征.而在归纳式推理数据集和全归纳式推理数据集上,有监督 SOTA 模型无法在训练阶段学习测试集上的实体表征,这限制了有监督 SOTA 模型的性能.而 RulePreM 不依赖于直接的实体和关系表征,而是基于规则来表征关系,因此在归纳式推理和全归纳式推理数据集上保持了良好的知识迁移能力.

#### 4.4 详细实验结果与分析

为了进一步分析 RulePreM 在不同数据集上的通用推理能力,分别报告了3种推理设定下每个数据集上的实验结果.

1)归纳式推理设定.图3显示了归纳式推理设定下的14个数据集上的详细  $MRR$  和  $Hits@10$  结果. RulePreM-预训练在12个数据集上优于同类型的 ULTRA-预训练模型; RulePreM-微调在11个数据集上优于同类型的 ULTRA-微调模型. RulePreM-预训练和 RulePreM-微调分别在10个和12个数据集上的结果优于有监督 SOTA 模型.这些观察验证了 RulePreM 基于规则增强推理在归纳式推理设定下的有效性.

2)全归纳式推理设定.图4显示了全归纳式推理设定下的13个数据集上的  $MRR$  和  $Hits@10$  结果. RulePreM-预训练模型在10个数据集上优于同类型的 ULTRA-预训练模型; RulePreM-微调在全部13个

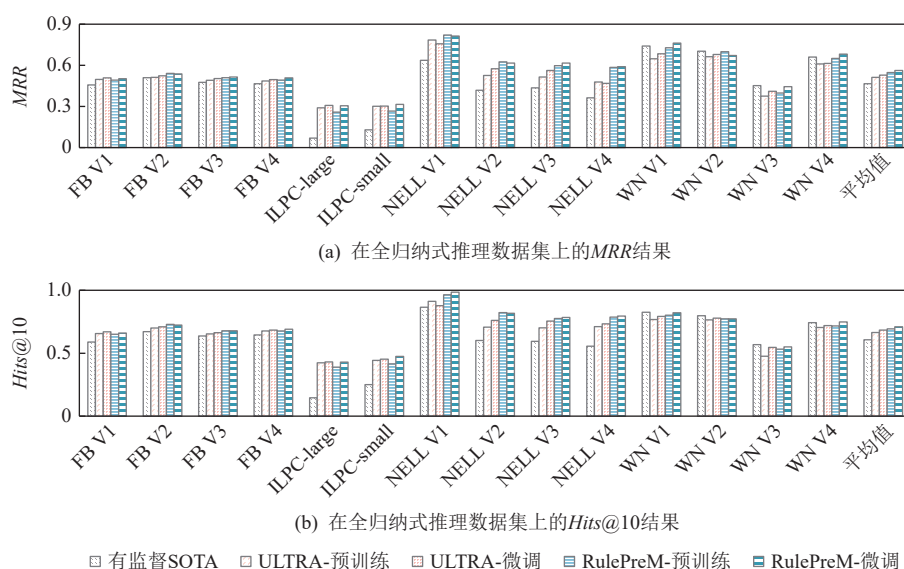


Fig. 3 The results of  $MRR$  and  $Hits@10$  on inductive reasoning datasets

图3 在归纳式推理数据集上的  $MRR$  与  $Hits@10$  结果

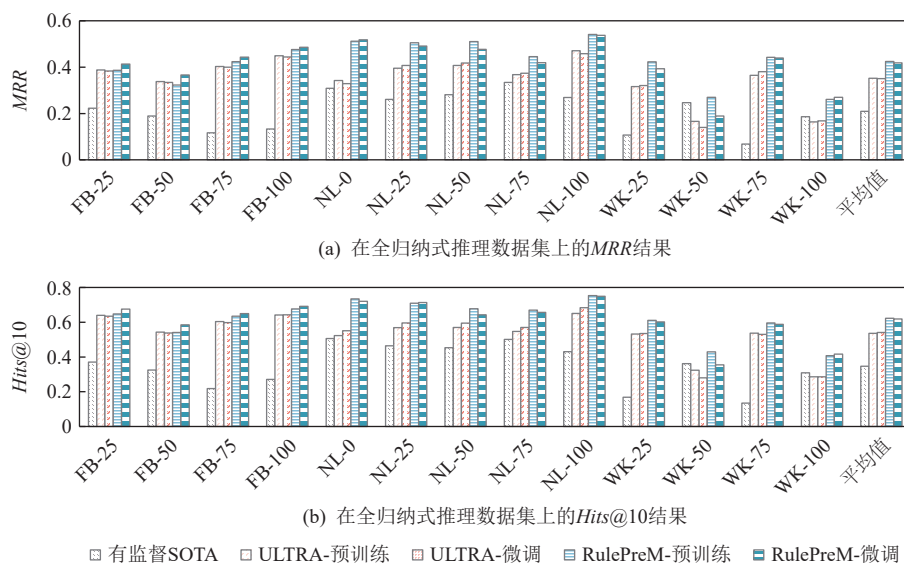


Fig. 4 The results of  $MRR$  and  $Hits@10$  on fully-inductive reasoning datasets

图4 在全归纳式推理数据集上的  $MRR$  与  $Hits@10$  结果

数据集上优于同类型的 ULTRA-微调模型. RulePreM-预训练和 RulePreM-微调在全部 13 个数据集上的结果均优于有监督 SOTA 模型, 而且与归纳式推理数据集上的结果相比, RulePreM 在全归纳式推理设定的性能提升更为明显.

3) 转导式推理设定. 图 5 显示了转导式推理设定

下的 16 个数据集上的  $MRR$  和  $Hits@10$  结果. 从实验结果中可以观察到, RulePreM-预训练和 RulePreM-微调的结果在 8 个数据集上均分别优于同类型方法 ULTRA-预训练和 ULTRA-微调. 与归纳式推理设定和全归纳式推理设定相比, RulePreM 在转导式推理上的优势并不明显.

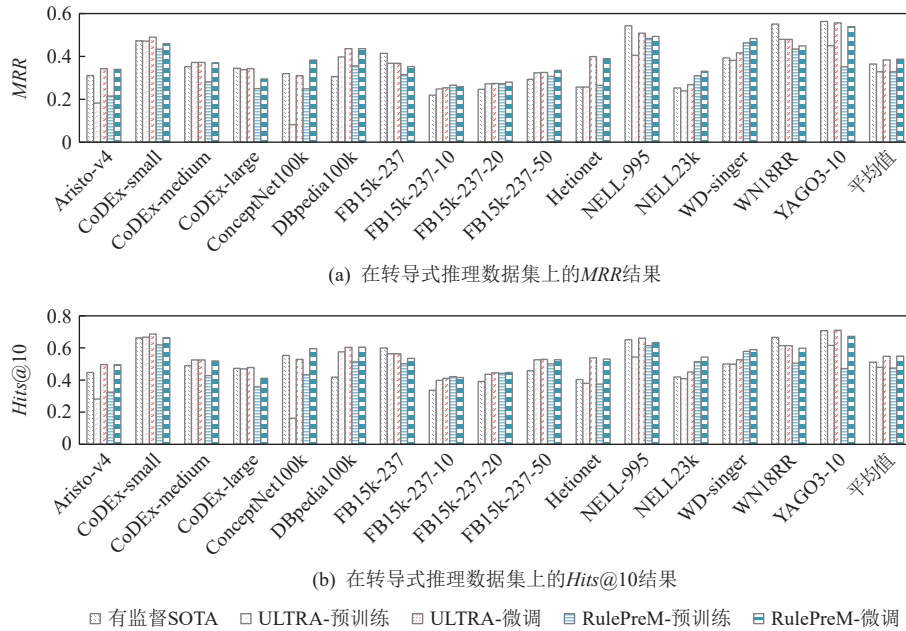


Fig. 5 The results of  $MRR$  and  $Hits@10$  on transductive reasoning datasets

图 5 在转导式推理数据集上的  $MRR$  与  $Hits@10$  结果

为了进一步分析在不同设定下的性能提升差异的原因, 在 4.7 节中报告了在不同推理设定下的规则数量, 并分析了数据集中的规则数量与质量对预训练模型性能的影响.

#### 4.5 消融实验

为了验证 RulePreM 中的关键技术对模型性能的影响与贡献, 设计了 RulePreM 的一些变体模型来移除某些模块, 并将其与完整的 RulePreM 模型做实验对比. 具体来说, 通过删除置信度阈值, 用所有规则来构建关系 IO 图, 构建了一个变体模型, 记作“w/o 置信度阈值”; 通过删除 RuleGNN 中的基于置信度的注意力, 构建了一个变体模型, 记作“w/o 置信度注意力”; 通过删除关系 IO 图构建与基于它的关系编码, 并用随机初始化的关系编码替代它, 构建了一个变体模型, 记作“w/o 关系 IO 图”.

表 2 展示了消融实验中完整的 RulePreM 与 2 个变体模型的对比. 可以观察到“w/o 置信度阈值”比完整模型有明显下降, 这表明了低质量推理模式对推理的负面影响; 相比于其它 2 个变体, “w/o 置信度注意力”变体模型的性能下降不是很明显, 这是因为置

信度阈值和置信度注意力机制都能起到过滤低质量规则的作用, 置信度阈值在关系 IO 图的构建环节已经过滤了大量低置信度的规则, 所以删除注意力模块后仍能保持一定的性能. 尽管如此, “w/o 置信度注意力”变体模型比完整模型在多项指标上仍有所下降, 这表明所提结合置信度的注意力机制能够在阈值过滤的基础上, 进一步降低在剩余规则中相对较低质量推理模式带来的负面影响, 增强知识图谱推理; “w/o 关系 IO 图”变体模型的结果与完整模型相比有明显下降, 这验证了所提关系 IO 图和 RuleGNN

Table 2 Results of Ablation Study

表 2 消融实验结果

| 模型          | 归纳式推理<br>(14 个数据集) |           | 全归纳式推理<br>(13 个数据集) |           | 转导式推理<br>(16 个数据集) |           | 平均值<br>(43 个数据集) |           |
|-------------|--------------------|-----------|---------------------|-----------|--------------------|-----------|------------------|-----------|
|             | $MRR$              | $Hits@10$ | $MRR$               | $Hits@10$ | $MRR$              | $Hits@10$ | $MRR$            | $Hits@10$ |
| 完整模型        | 0.550              | 0.698     | 0.420               | 0.621     | 0.331              | 0.476     | 0.429            | 0.592     |
| w/o 置信度阈值   | 0.380              | 0.522     | 0.299               | 0.459     | 0.162              | 0.259     | 0.275            | 0.405     |
| w/o 置信度注意力  | 0.522              | 0.697     | 0.419               | 0.620     | 0.323              | 0.473     | 0.417            | 0.590     |
| w/o 关系 IO 图 | 0.223              | 0.425     | 0.101               | 0.229     | 0.075              | 0.145     | 0.131            | 0.262     |

作为推理提示信息的有效性.

上述实验没有通过删除知识图谱的编码模块来构建变体模型, 因为编码模块是知识图谱推理模型的基础模块, 删除后预训练模型将无法运行. 因此, 增加了一个额外的实验来验证知识图谱编码模块的有效性. 具体来说, 采用 AnyBURL 100s 挖掘规则, 然后用所有挖掘得到的关系规则和混合规则执行推理并计算指标. 由于 AnyBURL 不能对所有实体排序, 也就不能提供准确的  $MRR$  分数, 所以在  $Hits@10$  指标上做了对比实验, 如图 6 所示. 可以观察到尽管 RulePreM 仅使用了关系规则, 它仍然在 3 类推理设定下均比 AnyBURL 取得了更优的性能. 一方面这是由于 AnyBURL 不能充分利用知识图谱中相关事实的信息; 另一方面是由于 RulePreM 能够在不同知识图谱之间迁移知识, 提高了推理性能.

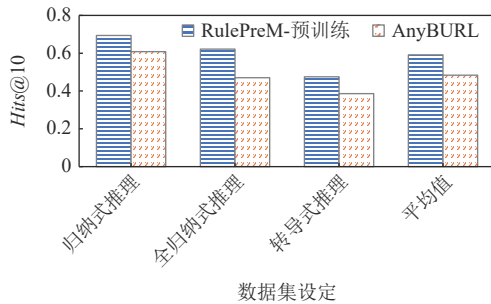


Fig. 6 Comparative experimental results with AnyBURL

图 6 与 AnyBURL 的对比实验结果

#### 4.6 关系编码器的变体实验

RulePreM 采用一个基于关系 IO 图的关系编码器 RuleGNN, 首先基于规则生成关系图, 然后再生成能够更好地建模方向信息的关系 IO 图, 最后基于多层消息传递和聚合来编码生成关系表示向量. 为了验证关系 IO 图和 RuleGNN 的有效性, 探索更多的关系编码方式, 设计了 2 种关系编码器的变体:

1) “关系图+RuleGNN”变体. 将关系 IO 图替换为简单的关系图, 并将式(1)改为所有的边向量均为零向量, 即  $\mathbf{r}^{(0)}=\mathbf{0}$ , 来表示关系图中无属性边; 然后将式(2)更改为:

$$\mathbf{v}^{(0)} = \begin{cases} \mathbf{q}, & \text{if } v = v_q; \\ \mathbf{0}, & \text{其他,} \end{cases} \quad (11)$$

其中  $\mathbf{q} \in \mathbb{R}^{1 \times d}$  表示一个随机初始化的一个可学习向量. 然后将关系图和更改后的初始化输入向量表示作为 RuleGNN 的输入, 得到关系的向量表示.

2) “规则体路径+LSTM”变体. 在抽取得到规则后, 在每条规则体前拼接规则头关系  $q$ , 并以一个可学习的随机初始化向量  $\mathbf{q} \in \mathbb{R}^{1 \times d}$  作为  $q$  关系的表示向

量, 其他关系被表示为零向量; 然后采用一个 3 层的 LSTM 神经网络分别编码每条规则的关系路径; 最后取每个关系表示向量的平均, 作为输出的关系表示向量来替代 RuleGNN 的输出.

表 3 显示了关系编码器的变体实验的结果, 其中“关系 IO 图+RuleGNN”表示 RulePreM 采用的关系 IO 图和 RuleGNN 的编码方式. 可以观察到将关系 IO 图替换为关系图之后, “关系图+RuleGNN”变体在 3 类设定的数据集上的结果均有所下降. 这表明所提关系 IO 图比简单的关系图能够更好地表征关系, 因为它更充分地利用了关系之间复杂的方向信息. 而“规则体路径+LSTM”变体的性能下降更加明显, 这表明因为仅对每条规则路径进行编码, 无法有效融合推理模式, 综合利用多条规则中的信息.

Table 3 Experimental Results of Variants of Rule-Based Relation Encoding

表 3 基于规则关系编码变体实验结果

| 模型              | 归纳式推理<br>(14 个数据集) |           | 全归纳式推理<br>(13 个数据集) |           | 转导式推理<br>(16 个数据集) |           | 平均值<br>(43 个数据集) |           |
|-----------------|--------------------|-----------|---------------------|-----------|--------------------|-----------|------------------|-----------|
|                 | $MRR$              | $Hits@10$ | $MRR$               | $Hits@10$ | $MRR$              | $Hits@10$ | $MRR$            | $Hits@10$ |
| 关系 IO 图+RuleGNN | 0.550              | 0.698     | 0.420               | 0.621     | 0.331              | 0.476     | 0.429            | 0.592     |
| 关系图+RuleGNN     | 0.500              | 0.682     | 0.392               | 0.582     | 0.278              | 0.439     | 0.385            | 0.561     |
| 规则体路径+LSTM      | 0.434              | 0.606     | 0.353               | 0.539     | 0.225              | 0.363     | 0.332            | 0.495     |

#### 4.7 规则置信度统计

为了深入探索 RulePreM 在不同设定下受规则挖掘的影响, 分别统计了归纳式推理、全归纳式推理和转导式推理设定下, 按照置信度阈值 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 筛选后, 平均每个关系包含的规则数量. 统计数据如图 7 所示, 从不同推理设定来看, 在所有规则置信度阈值下, 全归纳式推理数据集中每个关系的平均规则数最多, 其次是归纳式推理数据集, 而转导式推理数据集最少. 这解释了在主实验结果中, RulePreM 在全归纳式推理数据集上的性

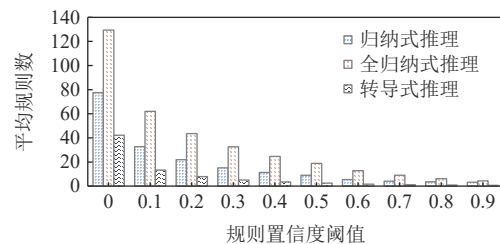


Fig. 7 Statistics of average number of rules per relation with various rule confidence threshold values

图 7 不同规则置信度阈值下平均每个关系的规则数统计



能提升最为明显,其次是归纳式推理数据集,而转导式推理数据集上的提升较小.转导式推理数据集上的规则数量太少,无法获取足够的信息来支持知识图谱推理,而另外2类设定下的规则较为充足,因此提升明显.另外,从规则置信度阈值来看,阈值设计有效过滤了大量低质量的推理规则.

#### 4.8 置信度阈值影响分析

为了进一步分析规则置信度阈值对于模型性能的影响,将规则置信度阈值分别设置为0, 0.1, 0.3, 0.5, 0.7来执行预训练和测试.实验结果如图8所示,在将规则置信度阈值设定为0时,RulePreM无法过滤低质量的推理模式,因此在3类推理设定中的性能均有明显下降.随着规则置信度阈值的提升,模型性能逐渐提升,并在0.3时达到最优性能.但当规则置信度阈值过高时,由于规则数量太少,关系IO图中包含的信息不足以有效支持知识图谱推理,模型性能也受到了影响.

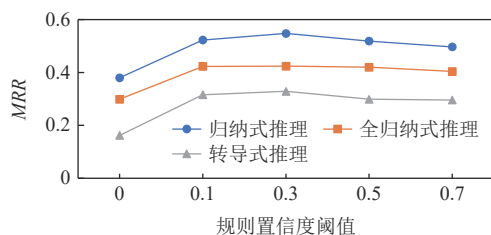


Fig. 8 Results of MRR under different rule confidence threshold values

图8 不同规则置信度阈值下的MRR结果

RulePreM旨在基于规则和置信度来筛选高质量路径,图6所示不同规则置信度阈值下的结果说明RulePreM能够有效筛选和利用高质量的推理模式,降低了低质量推理模式对推理造成的负面影响.综合上述结果与讨论,将规则置信度阈值设置在0.1~0.7,均能够有效降低低质量路径的负面影响,取得不错的推理性能.在实际应用中也可以从已知数据中抽取一部分数据作为验证集,并通过不同规则置信度阈值的预实验为具体数据集选择合适的规则置信度阈值.

## 5 总结与展望

受基于规则的方法启发,针对知识图谱推理模式的质量问题,提出了一种基于规则提示增强的知识图谱推理预训练模型RulePreM.具体来说,首先基于规则置信度来选择高质量推理规则,并构建关系IO图;然后提出一种规则关系编码器RuleGNN来编

码关系IO图来编码关系得到关系表示向量;最后基于知识图谱编码器得到实体表示向量,为候选实体打分和完成推理.另外,还提出一种结合规则置信度的注意力机制来减少低质量规则的影响.3种推理设定下的43个数据集上的实验结果表明,所提模型在不同知识图谱上具有良好的通用推理能力.

在规则置信度统计实验中,发现部分数据集上的高质量规则较少,限制了模型性能.因此在未来的研究工作中,打算继续研究如何设计规则不足的关系的编码来提高模型性能.另外,还计划研究如何将RulePreM部署在推荐系统和问答系统等下游应用中.

**作者贡献声明:**崔员宁提出了算法和完成实验;孙泽群提供论文写作指导和实验设计意见;胡伟提出指导意见并修改论文.

## 参 考 文 献

- [1] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review[J]. *Journal of Computer Research and Development*, 2016, 53(2): 247-261 (in Chinese)  
(刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. *计算机研究与发展*, 2016, 53(2): 247-261)
- [2] Liu Qiao, Li Yang, Duan Hong, et al. Knowledge graph construction techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600 (in Chinese)  
(刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582-600)
- [3] Ma Ang, Yu Yanhua, Yang Shengli, et al. Survey of knowledge graph based on reinforcement learning[J]. *Journal of Computer Research and Development*, 2022, 59(8): 1694-1722 (in Chinese)  
(马昂, 于艳华, 杨胜利, 等. 基于强化学习的知识图谱综述[J]. *计算机研究与发展*, 2022, 59(8): 1694-1722)
- [4] Wang Meng, Wang Haofen, Li Bohan, et al. Survey on key technologies of new generation knowledge graph[J]. *Journal of Computer Research and Development*, 2022, 59(9): 1947-1965 (in Chinese)  
(王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. *计算机研究与发展*, 2022, 59(9): 1947-1965)
- [5] Lehmann J, Isele R, Jakob M, et al. DBpedia-A large-scale, multilingual knowledge base extracted from Wikipedia[J]. *Semantic Web*, 2015, 6(2): 167-195
- [6] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase[J]. *Communications of the ACM*, 2014, 57(10): 78-85
- [7] Ji Shaoxiong, Pan Shirui, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(2): 494-514
- [8] Wang Meng, Wang Jingting, Jiang Yinlin, et al. Hybrid human-

- machine active search over knowledge graph[J]. *Journal of Computer Research and Development*, 2020, 57(12): 2501–2513 (in Chinese)  
(王萌, 王靖婷, 江胤霖, 等. 人机混合的知识图谱主动搜索[J]. *计算机研究与发展*, 2020, 57(12): 2501–2513)
- [9] Wang Quan, Mao Zhendong, Wang Bin, et al. Knowledge graph embedding: A survey of approaches and applications[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(12): 2724–2743
- [10] Guo Qingyu, Zhang Fuzhen, Qin Chuan, et al. A survey on knowledge graph-based recommender system[C]//Proc of the 39th IEEE Int conf on Data Engineering. Piscataway, NJ: IEEE, 2023: 3803–3804
- [11] Rossi A, Barbosa D, Firmani D, et al. Knowledge graph embedding for link prediction: A comparative analysis[J]. *ACM Transactions on Knowledge Discovery from Data*, 2021, 15(2): 1–49
- [12] Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners[C]//Proc of Annual Conf on Neural Information Processing Systems 2020. New York: Curran Associates, 2020: 1877–1901
- [13] Sun Zequn, Huang Jiacheng, Lin Jinghao, et al. Joint pre-training and local re-training: Transferable representation learning on multi-source knowledge graphs[C]//Proc of the 29th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2023: 2132–2144
- [14] Galkin M, Yuan Xinyu, Mostafa H, et al. Towards foundation models for knowledge graph reasoning[C]//Proc of 11th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2023: 1–14
- [15] Chen Mingyang, Zhang Wen, Geng Yuxia, et al. Generalizing to unseen elements: A survey on knowledge extrapolation for knowledge graphs[C]//Proc of the 32nd Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2023: 6574–6582
- [16] Hamaguchi T, Oiwa H, Shimbo M, et al. Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach[C]//Proc of the 26th Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2017: 1802–1808
- [17] Wang Peifeng, Han Jialong, Li Chenliang, et al. Logic attention based neighborhood aggregation for inductive knowledge graph embedding[C]//Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 7152–7159
- [18] Galkin M, Denis E G, Wu Jiapeng, et al. NodePiece: Compositional and parameter-efficient representations of large knowledge graphs[C]//Proc of the 10th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2022: 1–14
- [19] Teru K, Denis E, Hamilton W. Inductive relation prediction by subgraph reasoning[C]//Proc of the 37th Int Conf on Machine Learning. New York: PMLR, 2020: 9448–9457
- [20] Chen Jiajun, He Huarui, Wu Feng, et al. Topology-aware correlations between relations for inductive link prediction in knowledge graphs[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 6271–6278
- [21] Mai Sijie, Zheng Shuangjia, Yang Yuedong, et al. Communicative message passing for inductive relation reasoning[C]//Proc of the 35th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2021: 4294–4302
- [22] Xu Xiaohan, Zhang Peng, He Yongquan, et al. Subgraph neighboring relations infomax for inductive link prediction on knowledge graphs[C]//Proc of the 31st Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2022: 2341–2347
- [23] Zhu Zhaocheng, Zhang Zuobai, Xhonneux L P, et al. Neural Bellman-Ford networks: A general graph neural network framework for link prediction[C]//Proc of Annual Conf on Neural Information Processing Systems 2021. New York: Curran Associates, 2021: 29476–29490
- [24] Zhang Yongqi, Yao Quanming. Knowledge graph reasoning with relational digraph[C]//Proc of the 31st Int Conf on World Wide Web. New York: ACM, 2022: 912–924
- [25] Zhu Zhaocheng, Yuan Xinyu, Galkin M, et al. A\*Net: A scalable path-based reasoning approach for knowledge graphs[J]. arXiv preprint, arXiv: 2206.04798, 2022
- [26] Zhang Yongqi, Zhou Zhanke, Yao Quanming, et al. AdaProp: Learning adaptive propagation for graph neural network based knowledge graph reasoning[C]//Proc of the 29th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2023: 3446–3457
- [27] Lee J, Chung C, Whang J J. InGram: Inductive knowledge graph embedding via relation graphs[C]//Proc of the 40th Int Conf on Machine Learning. New York: PMLR, 2023: 18796–18809
- [28] Geng Yuxia, Chen Jiaoyan, Pan J Z, et al. Relational message passing for fully inductive knowledge graph completion[C]//Proc of the 39th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2023: 1221–1233
- [29] Gao Jianfei, Zhou Yangze, Ribeiro B. Double permutation equivariance for knowledge graph completion[J]. arXiv preprint, arXiv: 2302.01313, 2023.
- [30] Zhou Jincheng, Bevilacqua B, Ribeiro B. An OOD multi-task perspective for link prediction with new relation types and nodes[J]. arXiv preprint, arXiv: 2307.06046, 2023
- [31] Fang Yang, Zhao Xiang, Tan Zhen, et al. A Revised translation-based method for knowledge graph representation[J]. *Journal of Computer Research and Development*, 2018, 55(1): 139–150 (in Chinese)  
(方阳, 赵翔, 谭真, 等. 一种改进的基于翻译的知识图谱表示方法[J]. *计算机研究与发展*, 2018, 55(1): 139–150)
- [32] Yang Xiaohui, Wan Rui, Zhang Haibin, et al. Semantical symbol mapping embedding learning algorithm for knowledge graph[J]. *Journal of Computer Research and Development*, 2018, 55(8): 1773–1784 (in Chinese)  
(杨晓慧, 万睿, 张海滨, 等. 基于符号语义映射的知识图谱表示学习算法[J]. *计算机研究与发展*, 2018, 55(8): 1773–1784)
- [33] Dettmers T, Minervini P, Stenatorp P, et al. Convolutional 2D knowledge graph embeddings[C]//Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 1811–1818
- [34] Bordes A, Usunier N, García-Durán A, et al. Translating embeddings for modeling multi-relational data[C]//Proc of Annual Conf on Neural Information Processing Systems 2013. New York: Curran Associates, 2013: 2787–2795
- [35] Sun Zhiqing, Deng Zhihong, Nie Jianyun, et al. RotatE: Knowledge graph embedding by relational rotation in complex space[C]//Proc of the 7th Int Conf on Learning Representations. Washington DC:

- OpenReview. net, 2019: 1–18
- [36] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks[C]//Proc of the 8th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2020: 1–16
- [37] Galárraga L A, Teflioudi C, Hose K, et al. AMIE: Association rule mining under incomplete evidence in ontological knowledge bases[C]//Proc of the 22nd Int Conf on World Wide Web. New York: ACM, 2013: 413–422
- [38] Yang Fan, Yang Zhilin, Cohen W W. Differentiable learning of logical rules for knowledge base reasoning[C]//Proc of Annual Conf on Neural Information Processing Systems 2017. New York: Curran Associates, 2017: 2319–2328
- [39] Sadeghian A, Armandpour M, Ding P, et al. DRUM: End-to-End differentiable rule mining on knowledge graphs[C]//Proc of Annual Conf on Neural Information Processing Systems 2019. New York: Curran Associates, 2019: 15321–15331
- [40] Meilicke C, Chekol M W, Ruffinelli D, et al. Anytime bottom-up rule learning for knowledge graph completion[C]//Proc of the 28th Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2019: 3137–3143
- [41] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-Training of deep bidirectional transformers for language understanding[C]//Proc of the 2019 Conf the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2019: 4171–4186
- [42] Hou Zhenyu, Liu Xiao, Cen Yukuo, et al. GraphMAE: Self-supervised masked graph autoencoders[J]. arXiv preprint, arXiv: 2205.10803, 2022
- [43] Hu Weihua, Liu Bowen, Gomes J, et al. Strategies for pre-training graph neural networks[C]//Proc of the 8th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2020: 1–15
- [44] Rong Yu, Bian Yatao, Xu Tingyang, et al. Self-Supervised graph transformer on large-scale molecular data[C]//Proc of Annual Conf on Neural Information Processing Systems 2020. New York: Curran Associates, 2020: 1–13
- [45] Sun Fanyun, Hoffmann J, Verma V, et al. InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization[C]//Proc of the 8th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2020: 1–13
- [46] Velickovic P, Fedus W, Hamilton W L, et al. Deep graph infomax[C]//Proc of the 7th Int Conf on Learning Representations. Washington DC: OpenReview. net, 2019: 1–13
- [47] Fang Taoran, Zhang Yunchao, Yang Yang, et al. Prompt tuning for graph neural networks[J]. arXiv preprint, arXiv: 2209.15240, 2022
- [48] Liu Zemin, Yu Xingtong, Fang Yuan, et al. GraphPrompt: Unifying pre-training and downstream tasks for graph neural networks[C]//Proc of the 32nd Int Conf on World Wide Web. New York: ACM, 2023: 417–428
- [49] Gong Chenghua, Li Xiang, Yu Jianxiang, et al. Prompt tuning for multi-view graph contrastive learning[J]. arXiv preprint, arXiv: 2310.10362, 2023
- [50] Zhu Yun, Guo Jianhao, Tang Siliang. SGL-PT: A strong graph learner with graph prompt tuning[J]. arXiv preprint, arXiv: 2302.12449, 2023
- [51] Shirkavand R, Huang Heng. Deep prompt tuning for graph transformers[J]. arXiv preprint, arXiv: 2309.10131, 2023
- [52] Ma Yihong, Yan Ning, Li Jiayu, et al. HetGPT: Harnessing the power of prompt tuning in pre-trained heterogeneous graph neural networks[J]. arXiv preprint, arXiv: 2310.15318, 2023
- [53] Ge Qingqing, Zhao Zeyuan, Liu Yiding, et al. Enhancing graph neural networks with structure-based prompt[J]. arXiv preprint, arXiv: 2310.17394, 2023
- [54] Sun Xiangguo, Cheng Hong, Li Jia, et al. All in one: Multi-task prompting for graph neural networks[C]//Proc of the 29th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2023: 2120–2131
- [55] Chen Mouxiang, Liu Zemin, Liu Chenghao, et al. ULTRA-DP: ULTRA-unifying graph pre-training with multi-task graph dual prompt[J]. arXiv preprint, arXiv: 2310.14845, 2023
- [56] Sun Xiangguo, Zhang Jiawen, Wu Xixi, et al. Graph prompt learning: A comprehensive survey and beyond[J]. arXiv preprint, arXiv: 2311.16534, 2023
- [57] Chen Xuelu, Chen Muhao, Fan Changjun, et al. Multilingual knowledge graph completion via ensemble knowledge transfer[C]//Proc of the 2020 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 3227–3238
- [58] Singh H, Chakrabarti S, Jain P, et al. Multilingual knowledge graph completion with joint relation and entity alignment[C]//Proc of the 3rd Conf on Automated Knowledge Base Construction. Virtual: Online, 2021: 1–7
- [59] Huang Zijie, Li Zheng, Jiang Haoming, et al. Multilingual knowledge graph completion with self-supervised adaptive graph alignment[C]//Proc of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2022: 474–485
- [60] Chakrabarti S, Singh H, Lohiya S, et al. Joint completion and alignment of multilingual knowledge graphs[C]//Proc of the 2022 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2022: 11922–11938
- [61] Sun Zequn, Zhang Qingheng, Hu Wei, et al. A benchmarking study of embedding-based entity alignment for knowledge graphs[J]. Proceedings of the VLDB Endowment, 2020, 13(11): 2326–2340
- [62] Chen Muhao, Tian Yingtao, Yang Mohan, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]//Proc of the 26th Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2017: 1511–1517
- [63] Wu Yuting, Liu Xiao, Feng Yansong, et al. Relation-aware entity alignment for heterogeneous knowledge graphs[C]//Proc of the 28th Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2019: 5278–5284
- [64] Mao Xin, Wang Wenting, Wu Yuanbin, et al. Boosting the speed of entity alignment 10×: Dual attention matching network with normalized hard sample mining[C]//Proc of the 30th Int Conf on World Wide Web. New York: ACM, 2021: 821–832
- [65] Sun Zequn, Wang Chengming, Hu Wei, et al. Knowledge graph alignment network with gated multi-hop neighborhood aggregation[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 222–229



- [66] Sun Zequn, Hu Wei, Zhang Qingheng, et al. Bootstrapping entity alignment with knowledge graph embedding[C]//Proc of the 27th Int Joint Conf on Artificial Intelligence. Freiburg: IJCAI, 2018: 4396–4402
- [67] Wang Zhichun, Lv Qingsong, Lan Xiaohan, et al. Cross-lingual knowledge graph alignment via graph convolutional networks[C]//Proc of the 2018 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2018: 349–357
- [68] Zhang Wen, Zhu Yushan, Chen Mingyang, et al. Structure pretraining and prompt tuning for knowledge graph transfer[C]//Proc of the 32nd Int Conf on World Wide Web. New York: ACM, 2023: 2581–2590
- [69] Galkin M, Berrendorf M, Hoyt C T. An open challenge for inductive link prediction on knowledge graphs[J]. arXiv preprint, arXiv: 2203.01520, 2022
- [70] Toutanova K, Chen D. Observed versus latent features for knowledge base and text inference[C]//Proc of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. Stroudsburg, PA: ACL, 2015: 57–66
- [71] Xiong Wenhan, Hoang Thien, Wang W Y. DeepPath: A reinforcement learning method for knowledge graph reasoning[C]//Proc of the 2017 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 564–573
- [72] Mahdisoltani F, Biega J, Suchanek F M. YAGO3: A knowledge base from multilingual Wikipedias[C]//Proc of the 7th Biennial Conf on Innovative Data Systems Research (CIDR). [https://www.cidrdb.org/cidr2015/papers/CIDR15\\_paper1.pdf](https://www.cidrdb.org/cidr2015/papers/CIDR15_paper1.pdf)
- [73] Safavi T, Koutra D. CoDEX: A comprehensive knowledge graph completion benchmark[C]//Proc of the 2020 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 8328–8350
- [74] Lv Xin, Han Xu, Hou Lei, et al. Dynamic anticipation and completion for multi-hop reasoning over sparse knowledge graph[C]//Proc of the 2020 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 5694–5703
- [75] Chen Yihong, Minervini P, Riedel S, et al. Relation prediction as an auxiliary training objective for improving multi-relational graph representations[C]//Proc of the 3rd Conf on Automated Knowledge Base Construction. Virtual: Proceeding, 2021: 1–13
- [76] Ding Boyang, Wang Quan, Wang Bin, et al. Improving knowledge graph embedding using simple constraints[C]//Proc of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 110–121
- [77] Malaviya C, Bhagavatula C, Bosselut A, et al. Commonsense knowledge base completion with structural and semantic context[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 2925–2933
- [78] Himmelstein D S, Lizee A, Hessler C, et al. Systematic integration of biomedical knowledge prioritizes drugs for repurposing[J]. *eLife*, 2017, 6: e26726
- [79] Chen Yihong, Minervini P, Riedel S, et al. Relation prediction as an auxiliary training objective for improving multi-relational graph representations[C]//Proc of the 3rd Conf on Automated Knowledge Base Construction. Virtual: Proceeding, 2021: 1–13
- [80] He Tao, Liu Ming, Cao Yixin, et al. Exploring & exploiting high-order graph structure for sparse knowledge graph completion[J]. arXiv preprint, arXiv: 2306.17034, 2023
- [81] Guo Jia, Kok S. BiQUE: Biquaternionic embeddings of knowledge graphs[C]//Proc of the 2021 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2021: 8338–8351



**Cui Yuanning**, born in 1996. PhD candidate. His main research interests include knowledge graph, representation learning, and graph machine learning.

崔员宁, 1996年生. 博士研究生. 主要研究方向为知识图谱、表示学习、图机器学习.



**Sun Zequn**, born in 1992. PhD, Yuxiu Young Scholar. Member of CCF. His main research interests include knowledge graph, representation learning, and entity alignment.

孙泽群, 1992年生. 博士, 毓秀青年学者. CCF会员. 主要研究方向为知识图谱、表示学习、实体对齐.



**Hu Wei**, born in 1982. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include knowledge graph, database, and intelligent software.

胡伟, 1982年生, 博士, 教授, 博士生导师. CCF高级会员. 主要研究方向为知识图谱、数据库、智能软件.